OPEN ACCESS

Research Article

Towards sequential statistical testing as some standard: Pearson's correlation coefficient

Hin zu sequentiellem statistischem Testen als Standardverfahren: Pearsons Korrelationskoeffizient

Abstract

In standard statistical packages sequential tests are seldom and when they are offered, they mainly concern just tests about means. By a simulation study it was shown that a new sequential triangular test for the null-hypothesis H_0: $0<\rho\leq\rho_0$ for given requirements of precision (type-I-, type-II-risk, and a practical relevant effect $\delta=\rho_1-\rho_0$) offers reasonable results. Due to 100,000 runs the average sample size of the sequential triangular test is smaller than the sample size of the pertinent fixed sample size test. If practically possible, it is recommended that the triangular test should be used instead of a one-step procedure with a sample size fixed in advance.

Keywords: sequential test, correlation coefficient, simulation

Zusammenfassung

In statistischen Programmpaketen findet man kaum sequentielle Tests und wenn doch, werden vor allem Erwartungswerte getestet. Mit Hilfe einer Simulationsstudie konnte gezeigt werden, dass ein neuer sequentieller Dreieckstest der Nullhypothese H_0: $0<\rho\leq\rho_0$ bei gegebenen Genauigkeitsforderungen wie Risiko erster und zweiter Art und einem praktisch relevanten Effekt $\delta=\rho_1-\rho_0$) zu vernünftigen Ergebnissen führt. Bei 100.000 simulierten Tests war der durchschnittliche Stichprobenumfang des sequentiellen Dreieckstests kleiner als der entsprechende fest vorgegebene Umfang. Wenn es praktisch machbar ist, sollte stets der Dreieckstest an Stelle eines Tests mit fest vorgegebenem Umfang verwendet werden.

Schlüsselwörter: sequentieller Test, Korrelationskoeffizient, Simulation

Dieter Rasch[1]
Takuya Yanagida[2]
Klaus D. Kubinger[3]
Berthold Schneider[4]

1 Department of Landscape, Spatial and Infrastructure Sciences, University of Natural Resources and Life Sciences, Vienna, Austria

2 School of Applied Health and Social Sciences, University of Applied Sciences Upper Austria, Linz, Austria

3 Division of Psychological Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna, Vienna, Austria

4 Institute for Biometry, Hannover Medical School, Hannover, Germany

Introduction

In standard statistical packages sequential tests are seldom and when they are offered, they mainly concern just tests about means. For instance in [1], R-programs of sequential triangular tests can be found for:

- comparing a mean with a constant,
- comparing two means,
- comparing a probability with a constant,
- comparing two probabilities.

In this paper we present simulation results for a newly developed triangular sequential test for comparing a correlation coefficient with a constant.
In correlation analyses most of the time the null-hypotheses H_0: $\rho=0$ (versus the two-sided alternative hypothesis H_A: $\rho\neq0$) is tested. But a significant correlation coefficient – regardless of how small the type-I-risk α may be established and the more regardless how small the *p*-value results – has often no practical meaning. Therefore it is often more reasonable to test the null-hypothesis H_0: $\rho=\rho_0$ for any $0<\rho_0<1$ against the alternative H_A: $\rho=\rho_1$ for any $\rho_1>\rho_0$ or $\rho_1<\rho_0$. Kubinger et al. [2] made available even to SPSS-users the fixed sample size test, which is known for this problem since a long time in statistics. In the present paper we show the attractiveness of a corresponding sequential triangular test proposed by Schneider et al. [3]. In general, sequential triangular tests have the advantage that their average sample size is not only smaller than that of the corresponding fixed sample size tests but its maximum sample size needed is known in advance in any case.

Method

The null-hypothesis H_0: $\rho=\rho_0$ can – without loss of generality – be replaced by H_0: $0<\rho\leq\rho_0$ or by H_0: $\rho\geq\rho_0$. In an analogous way the alternative hypothesis may be replaced by a composite alternative.

Under the assumption that random variables are normally distributed with second moments σ^2_x, σ^2_y, and σ_{xy}, and a correlation coefficient $\rho=\sigma_{xy}/(\sigma_x\sigma_y)$ the null-hypothesis should be tested against the alternative hypothesis H_A: $0<\rho_0\leq\rho\leq\rho_1$ with a type-I-risk α – i.e. the probability of wrongly rejecting H_0 – and a type-II-risk β – i.e. the probability of wrongly accepting H_0 (in particular as long as $\delta\geq\rho_1-\rho_0$ with a ρ_1 to be fixed in advance).

The empirical correlation coefficient $r=s_{xy}/s_x s_y$ based on k observations (x_i,y_i) ($i=1,\ldots,k$) is an estimate of the parameter ρ (s_{xy}, s^2_x, s^2_y are the empirical covariance and variances, respectively). In place of using r as a test statistic we use the transformed (Fisher transformation [4]) value

$$z=\ln\left(\frac{1+r}{1-r}\right)$$

as a test statistic. The distribution of the corresponding random variable is approximately normal even if k is rather small. The expectation of z, being a function of ρ, amounts to

$$E(\mathbf{z})=\ln\left(\frac{1+\rho}{1-\rho}\right)+\frac{\rho}{n-1},$$

the variance to

$$\mathrm{var}(\mathbf{z})=\frac{4}{n-3}.$$

The statistic z can be used to test for a fixed sample size $k=n$ the hypothesis H_0: $\rho\leq\rho_0$ against the alternative hypothesis H_A: $\rho>\rho_0$ (respectively H_0: $\rho\geq\rho_0$ against H_1: $\rho<\rho_0$, a problem discussed in [5], Example 11.15). H_0: $\rho\leq\rho_0$ is rejected with error probability α, if

$$z\geq\ln\left(\frac{1+\rho}{1-\rho}\right)+\frac{\rho}{n-1}+z_{1-\alpha}\cdot\frac{2}{\sqrt{n-3}}$$

(respectively for H_0: $\rho\geq\rho_0$, if

$$z\geq\ln\left(\frac{1+\rho}{1-\rho}\right)+\frac{\rho}{n-1}-z_{1-\alpha}\cdot\frac{2}{\sqrt{n-3}};$$

$z_{1-\alpha}$ a is the (1-α)-quantile of the standard normal distribution).

An approximate lower bound for the sample size n which is necessary to keep a type-II-error with probability smaller or equal β given $\rho=\rho_1$ is the smallest positive integer for which holds:

$$n\geq 3+4\cdot\left(\frac{z_{1-\alpha}+z_{1-\beta}}{\ln(\frac{1+\rho_1}{1-\rho_1})-\ln(\frac{1+\rho_0}{1-\rho_0})}\right)^2.$$

A special group of sequential tests are the sequential triangular tests, going back to Whitehead [6] and Schneider [7]. The main characteristic is that based on the actual sampled date, to decide whether sampling has to be continued or either the null- or the alternative hypothesis can be accepted.

We split the sequence of data pairs into sub-samples of length, say $k>3$ each. For each sub-sample j ($j=1, 2,\ldots, m$) we calculate a statistic which distribution is only a function of ρ and k.

A triangular test must be based on a statistic with expectation 0, given the null-hypothesis, therefore we transform z into a realization of a standardized variable

$$z^*=\left[z-\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)-\frac{\rho_0}{k-1}\right]\frac{\sqrt{k-3}}{2} \quad (1)$$

which has the expectation 0 if the null-hypothesis is true. The parameter

$$\theta=\left[\ln\frac{1+\rho}{1-\rho}-\ln\frac{1+\rho_0}{1-\rho_0}+\frac{\rho-\rho_0}{k-1}\right]\frac{\sqrt{k-3}}{2} \quad (2)$$

is used as a test-parameter. For $\rho=\rho_0$ the parameter θ is 0 (as demanded). For $\rho=\rho_1$ we obtain:

$$\theta_1=\left[\ln\frac{1+\rho_1}{1-\rho_1}-\ln\frac{1+\rho_0}{1-\rho_0}+\frac{\rho_1-\rho_0}{k-1}\right]\frac{\sqrt{k-3}}{2} \quad (3)$$

The difference $\delta=\rho_1-\rho_0$ is the practical relevant difference which should be detected with the power $1-\beta$.

From each sub-sample j we now calculate the sample correlation coefficient r_j as well as its transformed values

$$z_j^*=\left[z_j-\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)-\frac{\rho_0}{k-1}\right]\frac{\sqrt{k-3}}{2} \quad (j=1, 2,\ldots, m).$$

Now by

$$Z_{m\theta=0}=\sum_{j=1}^{m}z^*_j$$

and $V_m=m$ the sequential path is defined by points (V_m, Z_m) for $m=1,2,\ldots$ up to the maximum of V below or exactly at the point where a decision can be made. The continuation region is a triangle whose three sides depend on α, β, and θ_1 via

$$a=\frac{\left(1+\frac{z_{1-\beta}}{z_{1-\alpha}}\right)\ln\left(\frac{1}{2\alpha}\right)}{\theta_1} \quad \text{and} \quad c=\frac{\theta_1}{2\left(1+\frac{z_{1-\beta}}{z_{1-\alpha}}\right)} \quad (4),$$

with the percentiles z_p of the standard normal distribution. That is, one side of the looked-for triangle lies between $-a$ and a on the ordinate of the (V, Z) plane ($V=0$). The two other borderlines are defined by the lines L_1: $Z=a+cV$ and L_2: $Z=-a+3cV$, which intersect at

$$\left(V_{\max}=\frac{a}{c}, Z_{\max}=2a\right) \quad (5).$$

The maximum sample size is of course $k\cdot V_{\max}$. The decision rule now is: Continue sampling as long as $-a+3cV_m<Z_m<a+cV_m$ if $\theta_1>0$ or $-a+3cV_m>Z_m>a+cV_m$ if $\theta_1<0$. Given $\theta_1>0$, accept H_A in case that Z_m reaches or exceeds L_1 and accept H_0 in case that Z_m reaches or underruns L_2, Given $\theta_1<0$, accept H_A in the case Z_m reaches or underruns L_1 and accept H_0 in the case Z_m reaches or exceeds L_2. If

the point $(V_{\max}=\frac{a}{c}, Z_{\max}=2a)$ is reached, H_A is to be accepted.

Example 1

We like to test the null hypothesis $\rho \leq .6$ against the alternative hypothesis $\rho > .6$ with $\alpha=.05$, $\beta=.1$ and $\delta=\rho_1-.6=.2$. We use $k=7$. We hence obtain

$$\ln\left(\frac{1+0.6}{1-0.6}\right)+\frac{0.6}{6}=1.486$$

and

$$\ln\left(\frac{1+0.8}{1-0.8}\right)+\frac{0.8}{6}=2.331\ .$$

Since $\sqrt{k-3}=2$ we obtain

$$\theta_1=\frac{2}{2}(2.331-1.486)=0.845\ .$$

We find $z_{.0.9}=1.282$ and $z_{.0.95}=1.645$, and hence

$$a=\frac{\left(1+\frac{1.282}{1.645}\right)\ln\left(\frac{1}{0.1}\right)}{0.845}=4.849$$

$$c=\frac{0.845}{2\left(1+\frac{1.282}{1.645}\right)}=0.168\ .$$

From (5) we get $V_{max}=4.849/0.168=28.863$, $Z_{max}=9.698$. The corresponding triangle is shown in Figure 1.

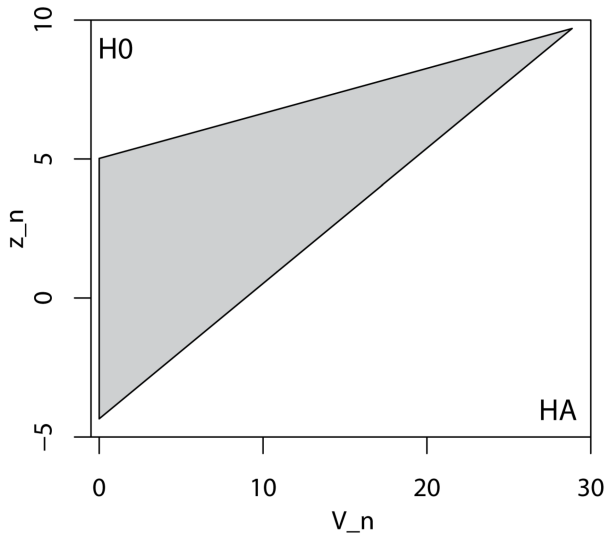

Figure 1: Graph of the triangle of example 1

Example 2

Finally we demonstrate the case of Example 11.15 in [5]. We like to test the null hypothesis $\rho \geq .8$ against the alternative hypothesis $\rho < .6$ with $\alpha=.05$, $\beta=.1$ and $\delta=.8-.6=.2$. We use $k=7$. We hence obtain

$$\ln\left(\frac{1+0.8}{1-0.8}\right)+\frac{0.8}{6}=2.331$$

and

$$\ln\left(\frac{1+0.6}{1-0.6}\right)+\frac{0.6}{6}=1.486\ .$$

Since $\sqrt{k-3}=2$ we obtain

$$\theta_1=\frac{2}{2}(1.486-2.331)=-0.845\ .$$

We find $z_{.0.9}=1.282$ and $z_{.0.95}=1.645$, and hence

$$a=\frac{\left(1+\frac{1.282}{1.645}\right)\ln\left(\frac{1}{0.1}\right)}{-0.845}=-4.849$$

$$c=\frac{-0.845}{2\left(1+\frac{1.282}{1.645}\right)}=-0.168\ .$$

From (5) we get $V_{max}=-4.849/-0.168=28.863$, $Z_{max}=-9.698$.
The corresponding triangle is shown in Figure 2.

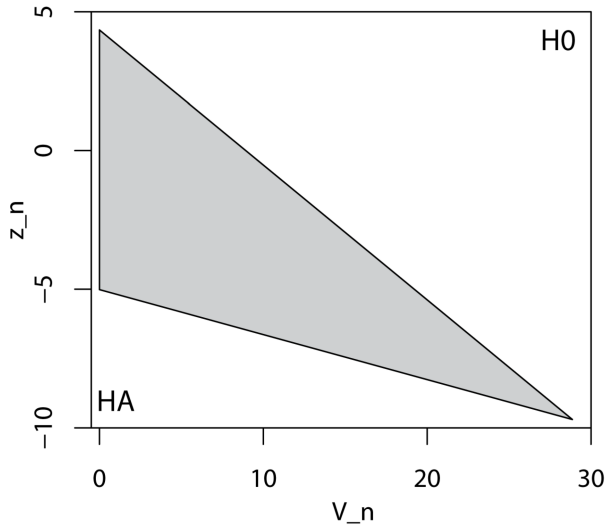

Figure 2: Graph of the triangle of example 2

Simulation study

The reasonability of the sequential triangular test for hypotheses of the correlation coefficient was tested by simulated paths (Z, V) being generated by bivariate normally distributed random numbers x and y with means $\mu_x=\mu_y=0$, variances $\sigma_x^2=\sigma_y^2=1$, and a correlation coefficient $\sigma_{xy}=\rho$. Simulations were performed with the nominal risks $\alpha_{nom}=.05$, $\beta_{nom}=.2$ and several values of ρ_0, ρ_1 and k. For each parameter combination 100,000 paths were generated. Simulations as well as the calculation of the introduced sequential triangular test were done in R [8]. As criteria for the test-quality we calculated: the relative frequency of wrongly accepting H_1, given $\rho=\rho_0$, which is the empirical risk of the first kind, say α_{emp}; the relative frequency of rejecting H_1, given $\rho=\rho_1$ – this is empirical

Table 1: Simulation results for α=.05 and β=.2 and 100,000 runs for each empty

ρ_0=.5, ρ_1=.6

k	12	16	20	50
α_{emp}	.0962	.0737	.0644	.0369
β_{emp}	.0815	.1052	.123	.1471
$ASN(\rho_0)$	249.0228	230.0896	224.274	221.09
$ASN(\rho_1)$	243.5088	244.2192	243.954	259.355
n_{fix}	302	302	302	302

ρ_0=.5, ρ_1=.65

k	12	16	20	50
α_{emp}	.0706	.054	.0464	.0336
β_{emp}	.0943	.1226	.1281	.1405
$ASN(\rho_0)$	104.2416	98.7344	96.778	104.62
$ASN\ \rho_1)$	111.6108	110.4512	111.126	125.71
n_{fix}	125	125	125	125

ρ_0=.5, ρ_1=.7

k	12	16	20	50
α_{emp}	.0531	.042	.0413	.027
β_{emp}	.1135	.1304	.1329	.1378
$ASN(\rho_0)$	55.6608	54.4528	54.29	64.275
$ASN\ \rho_1)$	62.0748	62.3424	63.024	75.405
n_{fix}	65	65	65	65

ρ_0=.6, ρ_1=.7

k	12	16	20	50
α_{emp}	.0937	.0747	.0643	.0373
β_{emp}	.0781	.1018	.1153	.1408
$ASN(\rho_0)$	176.712	163.9968	159.78	159.675
$ASN\ \rho_1)$	173.8512	172.2336	173.034	188.94
n_{fix}	207	207	207	207

ρ_0=.6, ρ_1=.75

k	12	16	20	50
α_{emp}	.0632	.0517	.0454	.0249
β_{emp}	.1023	.1195	.1256	.1348
$ASN(\rho_0)$	71.3652	67.8848	67.58	76.035
$ASN\ \rho_1)$	77.034	76.1136	77.952	90.855
n_{fix}	82	82	82	82

ρ_0=.6, ρ_1=.8

k	8	12	16	20	50
α_{emp}	.0667	.048	.038	.033	.0207
β_{emp}	.0803	.0885	.1164	.1234	.1034
$ASN(\rho_0)$	41.3096	37.0776	36.9648	37.364	52.455
$ASN(\rho_1)$	43.2424	41.5812	42.7376	43.986	57.88
n_{fix}	41	41	41	41	41

ρ_0=.7, ρ_1=.8

k	16	20	50
α_{emp}	.064	.057	.0324
β_{emp}	.1015	.1117	.1367
$ASN(\rho_0)$	98.1936	96.104	102.67
$ASN(\rho_1)$	104.9072	105.5	120.785
n_{fix}	119	119	119

ρ_0=.7, ρ_1=.85

k	12	16	20	50
α_{emp}	.0495	.0406	.0348	.0233
β_{emp}	.0976	.1141	.1267	.1021
$ASN(\rho_0)$	40.4028	40.1568	40.37	53.985
$ASN(\rho_1)$	44.5476	45.4208	46.676	60.495
n_{fix}	44	44	44	44

ρ_0=.7, ρ_1=.9.

k	8	12	16	20
α_{emp}	.0466	.038	.0288	.0282
β_{emp}	.0861	.1028	.106	.0943
$ASN(\rho_0)$	21.2232	20.1612	20.7248	22.54
$ASN(\rho_1)$	23.072	22.8144	23.7568	25.718
n_{fix}	20	20	20	20

ρ_0=.8, ρ_1=.9.

k	12	16	20	50
α_{emp}	.0592	.0466	.0433	.0233
β_{emp}	.0943	.1097	.1198	.0975
$ASN(\rho_0)$	44.7372	43.2608	42.996	55.49
$ASN(\rho_1)$	47.8344	48.352	49.404	62.72
n_{fix}	48	48	48	48

ρ_0=.8, ρ_1=.95.

k	8	12	16	20
α_{emp}	.0461	.0306	.0263	.0241
β_{emp}	.0799	.1008	.0825	.1078
$ASN(\rho_0)$	15.8192	15.414	17.312	20.354
$ASN(\rho_1)$	17.4848	17.5512	19.1984	21.11
n_{fix}	15	15	15	15

risk of the second kind, say β_{emp}; the average number of samples *ASN* for ρ_0 and ρ_1 used for the calculation of r and z^* until data sampling stopped (that is: the path leaves the continuation region). Here, *ASN* is the mean number of sample pairs over all 100,000 runs of the simulation study. Bear in mind that in a certain case the number of sample pairs needed for a terminal decision may lie either below or above that value *ASN*. The procedure is reasonable, if $\alpha_{emp} \leq \alpha$ and $\beta_{emp} \leq \beta$. Of course, the sequential test should also lead to $ASN < n_{fix}$, where n_{fix} is the sample size necessary in a fixed sample size test with a power $1-\beta$, given $\rho=\rho_1$ and testing the hypothesis $\rho \leq \rho_0$

with the type-I-risk α=.05. Results are shown in Table 1 where the difference between α_{emp} and α are not to large. The *ASN* strongly depends on the real value of ρ. From Table 1 the reader finds reasonable values of k for the pairs (ρ_0, ρ_1) to be tested.

To summarize: If one actually wishes to check whether a correlation coefficient is large enough we suggest first to fix a lower bound of the coefficient of determination ρ_0^2 as the value which at least must be given in order to speak of a meaningful percentage of the variance of y which can be explained by x – and vice versa. Another value of the coefficient of determination, ρ_1^2, is that one which the researcher must not oversee with a higher probability than 1–β if given. It is easy to adapt the procedure for negative roots of these squares. For example, one might decide for ρ_0^2=.5, as a consequence of which ρ_0=.7; suppose ρ_1=.8, α=.05, β=.2, then using k=26 (by interpolation) a total sample size will result as about 108 instead of 119 for a fixed sample study.

Note

Conflict of interest

The authors declare that they have no competing interests.

References

1. Rasch D, Pilz J, Verdooren RL, Gebhardt A. Optimal experimental design with R. New York: Chapman & Hall/CRC; 2011.
2. Kubinger KD, Rasch D, Šimeckova M. Testing a correlation coefficient's significance: Using H0: 0<ρ≤λ is preferable to H0: ρ=0. Psychol Sci. 2007;49:74-87.
3. Schneider B, Rasch D, Kubinger KD, Yanagida T. A Sequential Triangular Test of a Correlation Coefficient's Null-Hypothesis: 0<ρ≤ρ0. Stat Pap. 2014 Jun 12. DOI: 10.1007/s00362-014-0604-8
4. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika. 1915;10(4):507-21.
5. Rasch D, Kubinger KD, Yanagida T. Statistics in Psychology. Using R and SPSS. Chichester: Wiley; 2011. DOI: 10.1002/9781119979630
6. Whitehead J. The Design and Analysis of Sequential Clinical Trials. 2nd ed. Chichester: Ellis Horwood; 1992.
7. Schneider B. An interactive computer program for design and monitoring of sequential clinical trials. In: Proceedings of the XVIth international biometric conference; 1992 Dec 7-11; Hamilton, New Zealand. p. 237-50.
8. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available from: http://www.R-project.org/

Corresponding author:
Dieter Rasch
Department of Landscape, Spatial and Infrastructure Sciences, University of Natural Resources and Life Sciences, Peter-Jordan-Straße 82, 1190 Wien, Austria
d_rasch@t-online.de

Please cite as
Rasch D, Yanagida T, Kubinger KD, Schneider B. Towards sequential statistical testing as some standard: Pearson's correlation coefficient. GMS Med Inform Biom Epidemiol. 2014;10(1):Doc07. DOI: 10.3205/mibe000156, URN: urn:nbn:de:0183-mibe0001569

This article is freely available from
http://www.egms.de/en/journals/mibe/2014-10/mibe000156.shtml

Published: *2014-10-28*